

Update on 2020 Census Detailed
Demographic and Housing Characteristics
File (Detailed DHC)
CSAC Discussant

Lance A. Waller, Ph.D.

Department of Biostatistics and Bioinformatics

Rollins School of Public Health

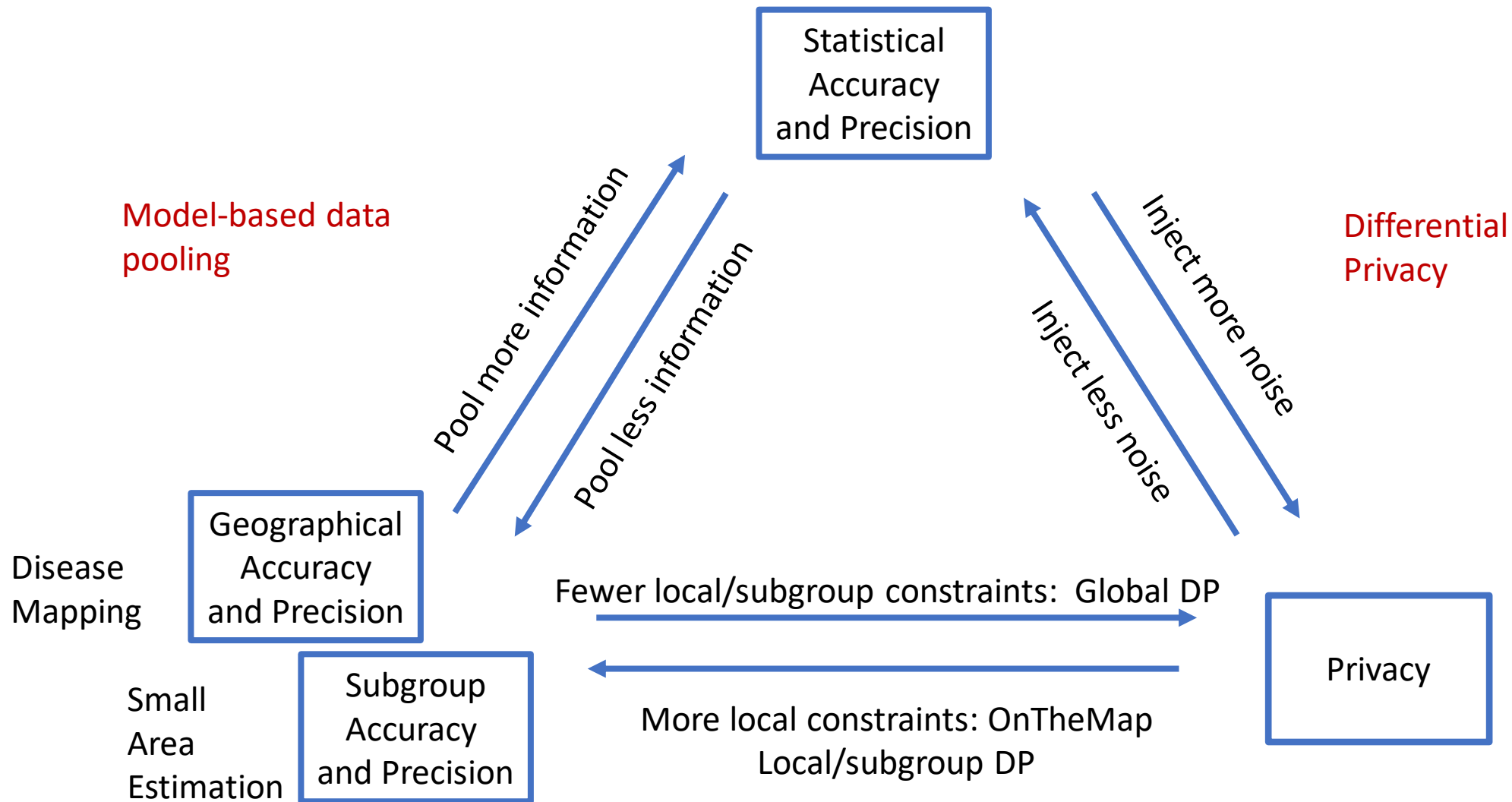
Emory University

Questions for the CSAC

- **Q1: What parameter settings (e.g., MOEs, levels of geography) are acceptable to meet data user needs?**
- **Q2: Do you have any feedback on the Tumult Labs methodological work for developing tables T1 and T2?**
- **Q3: Do you have any feedback on the methodology to model Join tables?**

Summary of discussion

- Thank you for the hard work so far.
- Limited time to respond to a lot of information.
 - Initial thoughts only
 - Recommend follow-ups with CSAC Differential Privacy Working Group
- Setting: Three sources of analytical tension.
- Q1 and Q2: Differential privacy and proposed 2020 Detailed DHC data.
- Q3: Statistical modeling and join tables.



Detailed DHC:
What to provide and how to provide it?

Data Product Comparisons: 2010 vs 2020

Topics	2010	2020
Number of Population Tables	57	4
Number of Geographies	104	4
Threshold for National Sex by Age Characteristics	100	500
Threshold for National Total Count for <ul style="list-style-type: none">• AIAN Tribes and Villages• Detailed Race and Ethnicity Groups	0 100	10 10
Number of Race and Ethnicity Groups (not inclusive of AIAN tribes and villages)	67	334
Disclosure Avoidance Technique	Swapping	Differential Privacy

Note: Estimates based on 2010 Census SF-2 data, 2016 ACS 5-Year data, and 2020 Census proposal for the Detailed DHC.

Comments

- Large changes in numbers of available tables
 - Not just reduction in numbers of tables, differences are more subtle.
 - Will require careful communication.
 - Response of “Differential Privacy” will not be reason enough.

Q1: What parameter settings (e.g., MOEs, levels of geography) are acceptable to meet data user needs?

- Levels of Geography:
 - Proposed release of Detailed DHC for Nation; State; County; American Indian, Alaska Native, and Native Hawaiian (AIANNH) tribes and villages.
 - Users will want tract-level values for local decisions.
 - Some counties have smaller population sizes than typical tracts (2019 ACS)
 - Small-population counts are not uniformly spread in space (more in rural areas, more in West)
 - Difficult question: Same level of aggregation everywhere?
- Margins of error:
 - Important feature (greater MOE, less need for additional noise)
 - Will need use-case specific discussions

Q2: Do you have any feedback on the Tumult Labs methodological work for developing tables T1 and T2?

- Plans for Tables T1 and T2 (individual demographics)
- Plans for Tables T3 and T4 (household demographics)
- Tumult document provides technical details.
- List of seven desired features helpful:
 - Privacy, Population Groups, **Adaptivity**, Accuracy (MOE), Integrality, **No Consistency** (adjusted in separate algorithm).
- Adaptivity: Different levels of aggregation for different population groups.
- Communication of differential privacy
 - Technical parameters (epsilon, delta, MOE): convert to (comparative) risk of exposure/discovery?

Q3: Do you have any feedback on the methodology to model Join tables?

- Modeling
 - Long history of small area modeling by/supported by Census Bureau
 - Fay and Herriot, Malec, Datta, Holan, others.
 - Compromise between design-based and model-based statistical reporting.
 - Design-based: Differential privacy literature
 - Model-based: Synthetic data literature
- CSAC Differential Privacy Working Group would benefit from follow up(s) with modeling group.

Q3 response continued

- Not only for join tables but potential data reporting at tract level.
 - Challenging but offers mechanism for broader, model-based releases.
 - Need further discussion contrasting :
 - “Synthetic Data” (sampled from posterior)
 - Posterior mean predictions
 - Differentially private counts
- The “market” for tract-level data will drive creation of data products.
Where will the Census Bureau’s data releases fit in?

Summary

- Clear progress and a lot of work.
- Further discussions needed:
 - Linking use case classes to MOE and geographies
 - Discussing options for tract-level data (expand modeling efforts)
 - Synthetic data
 - Differentially private data
 - Hybrid?
 - Framing output products of modeling:
 - Official data releases
 - Methods available for customized data products
 - Communicating data products to multiple constituencies